



The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness

RAND RESEARCH AREAS
CHILDREN AND ADOLESCENTS
CIVIL JUSTICE
EDUCATION
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INTERNATIONAL AFFAIRS
U.S. NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
SUBSTANCE ABUSE
TERRORISM AND
HOMELAND SECURITY
TRANSPORTATION AND
INFRASTRUCTURE

20041008 335

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of individual published, peer-reviewed documents or of a body of published work.

Corporate Headquarters
1700 Main Street
P.O. Box 2138
Santa Monica, California
90407-2138
TEL 310.393.0411
FAX 310.393.4818

© RAND 2004

www.rand.org

Value-added modeling (VAM), a collection of statistical techniques that uses multiple years of student test score data to estimate the effects of individual schools or teachers, has recently garnered a great deal of attention among both policymakers and researchers. For example, several states—including Tennessee, Pennsylvania and Ohio—are providing at least some of their schools and school districts with feedback about their performance based on VAM, and, in some statehouses, the idea of using VAM results to evaluate and reward administrators and teachers has been discussed.

This interest on the part of policymakers reflects the promise of VAM, but many technical issues must be considered in the execution and application of VAM to ensure that policy decisions are based on sound information. Although there have been reviews of particular approaches, no previous reviews carefully compared recent VAM efforts or systematically discussed the wide variety of issues they raise. To address this problem, RAND researchers, funded by the Carnegie Corporation of New York, undertook a systematic review and evaluation of leading approaches to VAM. The goals of this investigation were to

- delineate the technical issues raised by the use of VAM for measuring teacher performance
- evaluate the practical impact of decisions regarding modeling techniques, variations in the quality of the data used in modeling processes, choices of outcome measures, and techniques for sampling student performance
- identify gaps in the literature that could benefit from further research
- inform the debate among both researchers and policymakers about the potential of VAM.

In addition, the research team estimated the effects of math teachers for students in Grades 3–5,

Abstract

Value-added modeling offers the possibility of estimating the effects of teachers and schools on student performance, a potentially important contribution in the current environment of concern for accountability in education. These techniques, however, are susceptible to a number of sources of bias, depending on decisions about how the modeling is executed and on the quality of the data on which models are based. If teachers are to be held accountable for the performance of their students, strategies for measuring the impact of their work must be refined or, at least, the uncertainties of these measurements must be taken into account in assessing the impact of teachers and schools on student performance.

using math scores from a sample of schools in a large suburban district. This independent analysis permitted examination of the effects of certain variations in modeling strategies.

Value-Added Modeling Has the Potential to Identify Effects of Teachers on Student Performance

VAM attempts to determine the incremental effects of inputs into education, controlling for the prior achievement level of students. In practice, VAM is used to estimate the unique contributions of the school or teacher on students' progress over the course of a year rather than the cumulative effects of education or student background factors.

Two factors have contributed to recent interest in VAM. First, in theory, VAM has the potential to

separate the effects of teachers and schools on student performance from the powerful effects of noneducational factors such as family background. This isolation of the effects of educational and noneducational factors is critical for accurate evaluation of schools and teachers. Second, some recent VAM studies purport to show very large differences in effectiveness among teachers. If these differences can be substantiated and can be causally linked to specific characteristics of teachers, significant improvements in education could be made through the selection of effective teachers or through training to improve teacher effectiveness.

Variations in Teachers Affect Student Performance, but Size of Effect Is Uncertain

The recent literature on VAM suggests that teacher effects on student learning are large, accounting for a significant portion of the variability in growth, and that they persist for at least three to four years into the future. RAND researchers critically evaluated the methods used in these studies and the validity of the resulting claims. They concluded that teachers do, indeed, have discernible effects on student achievement and that these teacher effects appear to persist across years.

The shortcomings of existing studies, however, make it difficult to determine the size of teacher effects. Nonetheless, it appears that the magnitude of some of the effects reported in these studies is overstated. To determine the true size of teacher effects, several important statistical and psychometric issues must be addressed.

We group these issues into four categories: basic issues of statistical modeling; issues involving omitted variables, confounders, and missing data; issues arising from the use of achievement test scores as dependent measures; and uncertainty about estimated effects.

Impact of Alternative Statistical Modeling Strategies on Estimates of Teacher Effects

Modeling choices could have a significant impact on estimates of teacher performance. The problem of small classes is a case in point. When the number of students taught by a particular teacher is small, estimates of teacher effects can be heavily influenced by the performance of only a few students. One modeling approach to addressing this problem involves using data from small classes without adjusting for class size. This approach, however, tends to classify too many teachers of small classes as either highly effective or highly ineffective. An alternative approach, used in many of the most prominent recent VAM studies, "shrinks" estimates for individual teachers back toward the overall mean. That is, estimates of the effects of teachers who teach small numbers of students are statistically adjusted so that they are similar to the average effect of all teachers. This approach offsets the problem of distortions in the overall effects of teachers, but it makes identifying particularly effective or ineffective teachers who teach small classes considerably more difficult.

Impact of Omitted Variables, Confounders, and Missing Data on Estimates of Teacher Effects

In VAM, analysts rely on observational, rather than experimental, data. Reliance on such data can lead to inaccuracy in estimates of

teacher effects due to (1) differences between schools or classrooms that are not fully controlled in the analysis (such differences "confound" the results) and (2) shortcomings of the data collected within schools.

Impact of Absence of Controlled Comparisons Across Schools.

When differences between schools are not experimentally controlled, influences on student learning by factors other than teachers, such as other characteristics of the school in which the teacher works, may not be properly accounted for. For instance, if students attending different schools differ in ways that are likely to affect both achievement and growth in achievement and if the composition of the school's students (e.g., the proportion of students eligible for free and reduced-price lunches) affects these outcomes, bias in estimates of teacher effects can occur.

Some recent work on this topic suggests that variations in individual student characteristics have little influence on estimated teacher effects, but our own exploration suggests that the composition of the school had a great impact on estimates of teachers' effectiveness. We conducted a limited investigation of performance in mathematics—three grades in one school district were examined—and found that the composition of the school does affect growth in some settings. Thus, if variations in the composition of the school are not taken into account, these omitted variables may produce bias in applications of VAM. Because true teacher effects might be correlated with the characteristics of the students they teach, current VAM approaches cannot separate effects caused by the composition of the school from teacher effects.

Also difficult to disentangle from the effect of the students' current teachers are other characteristics of schools (i.e., characteristics other than the composition of the student body), of districts, or of prior teachers. If these variables are omitted from the analysis, their effects are subsumed by the estimated teacher effects. Alternatively, if such effects are included in models and if teachers of differing effectiveness cluster at the school or district level, part of the true teacher effects will be attributed to schools or to districts. Both approaches may result in biased estimation of the true teacher effects. Analysts must decide which potential error is more acceptable.

Impact of Missing Data. Longitudinal student achievement data will inevitably be incomplete. Information regarding the performance of individual students, as well as data linking students to teachers, may be lacking. Estimates of teacher effects may be sensitive to both the nature of missing data and the analytic approach used to address the problem. For example, if the test scores of low-performing students are missing, the scores of high-performing students will have a disproportionate impact on estimates of teacher effectiveness, possibly making teachers appear more effective than is, in fact, the case. Little is currently known about the effects of missing data on VAM estimates of teacher effects, but the potential for bias is large because the factors that contribute to missing links and missing test scores are common: Students are mobile, with large proportions transferring among schools every year.

Effects of Using Achievement Tests as an Outcome

VAM uses measures of student achievement to define and estimate

teacher effects, but these achievement measures are limited in several ways. Changes in the timing of tests, the weight given to alternative topics, or the methods used to create scores from students' responses (the "scaling" of the test) could affect conclusions about the relative achievement or growth in achievement across classes of students.

Such changes would, in turn, change estimates of teacher effects. In some cases, the effects could be substantial. For example, in a middle school in which curriculum is differentiated, a test emphasizing advanced content may favor teachers instructing the most able students, while a test emphasizing more basic content may boost the estimated impact of those teaching less advanced students.

Effects of Sampling Error

Sampling error is another potential source of error in VAM estimates. Estimates of teacher effects have larger sampling errors than estimates of school effects because of the smaller numbers of students used in the estimation of individual teacher effects. Thus, some estimates of interest will be too unreliable to use. Even so, for some purposes, such as identifying teachers who are extremely effective or ineffective, the estimates might be sufficiently precise. However, for other purposes, such as ranking teachers, the uncertainty in the estimates is likely to be too large to allow anything to be said with any degree of confidence.

Recommendations

Using VAM to estimate individual teacher effects is a recent endeavor, and many of the possible sources of error have not been thoroughly evaluated in the literature. The goal of this study was to identify possible sources of error and bias and evaluate what is

known at this point. To improve the quality and usefulness of VAM in the future, the authors recommend that researchers

- develop databases that can support VAM estimation of teacher effects across a diverse sample of school districts or other jurisdictions
- develop computational tools for fitting VAM that scale up to large databases and allow for extensions to the currently available models
- link estimates of teacher effects derived from VAM with other measures of teacher effectiveness as a means of validating estimate effects
- conduct further empirical investigation on the impact of potential sources of error in VAM estimates
- determine the prevalence of factors that contribute to the sensitivity of estimated teacher effects
- incorporate decision theory into VAM by working with policymakers to elicit decisions and costs associated with those decisions and by developing estimators to minimize the losses.

The Bottom Line

The current research base is insufficient to support the use of VAM for high-stakes decisions, and applications of VAM must be informed by an understanding of the potential sources of errors in teacher effects. Policymakers, practitioners, and VAM researchers need to work together so that research is informed by the practical needs and constraints facing users of VAM and so that implementation of the models is based on the kinds of inferences and decisions the research currently supports. If teachers are to be held accountable for the performance of their students, they deserve the best measurement of their effects on students that we can provide. ■

This research brief describes work done for RAND Education documented in *Evaluating Value-Added Models for Teacher Accountability* by Daniel F. McCaffrey, Daniel M. Koretz, J.R. Lockwood, and Laura S. Hamilton, MG-158-EDU, 2004, 154 pages, available online at www.rand.org/publications/MG/MG158/ and from RAND Distribution Services (phone: 310-451-7002; toll free: 877-584-8642; or email: order@rand.org). The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

RAND Offices Santa Monica • Washington • Pittsburgh • New York • Doha • Berlin • Cambridge • Leiden